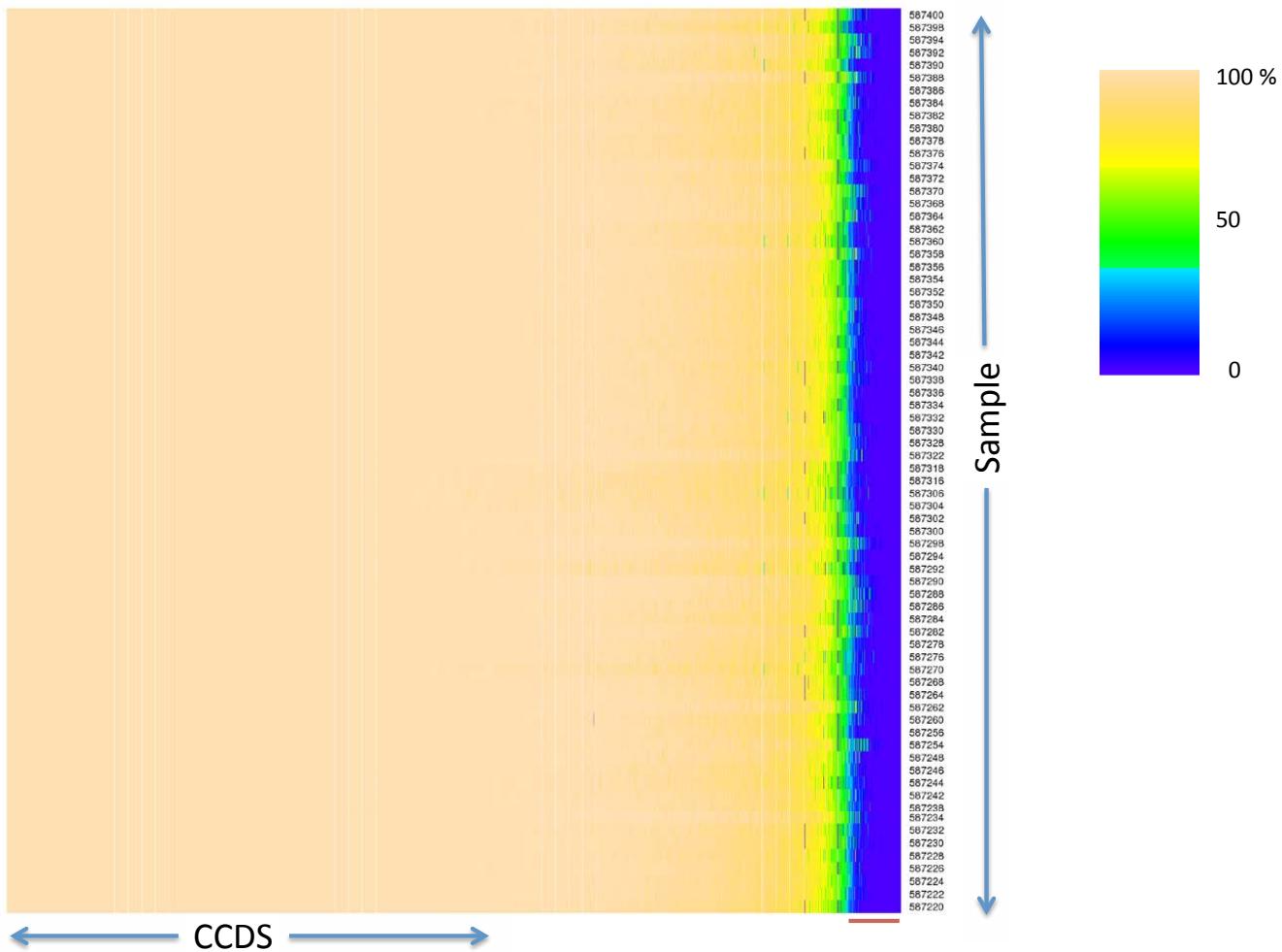
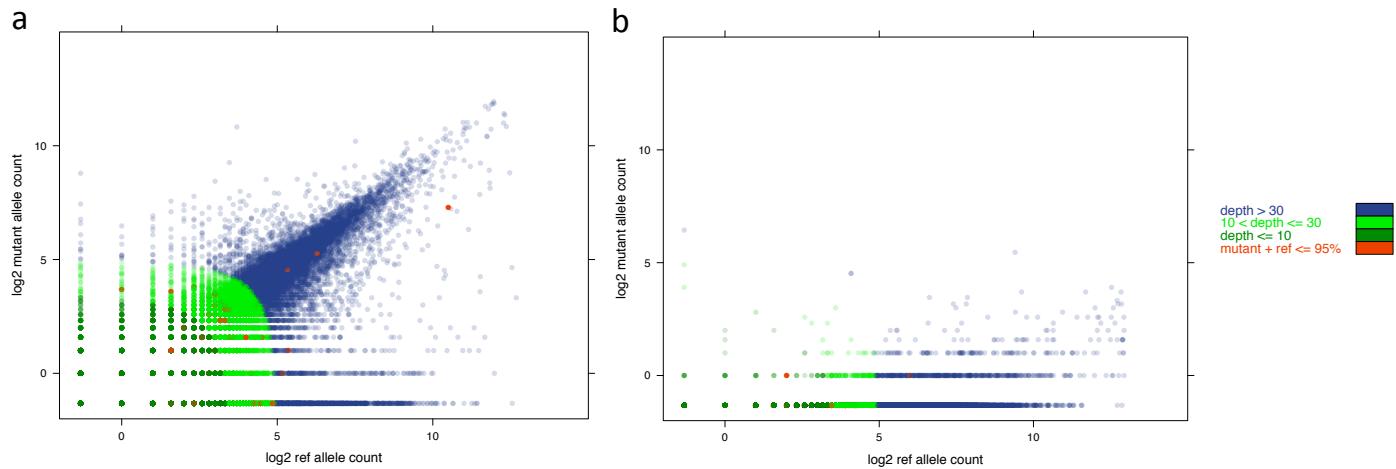


SUPPLEMENTARY INFORMATION

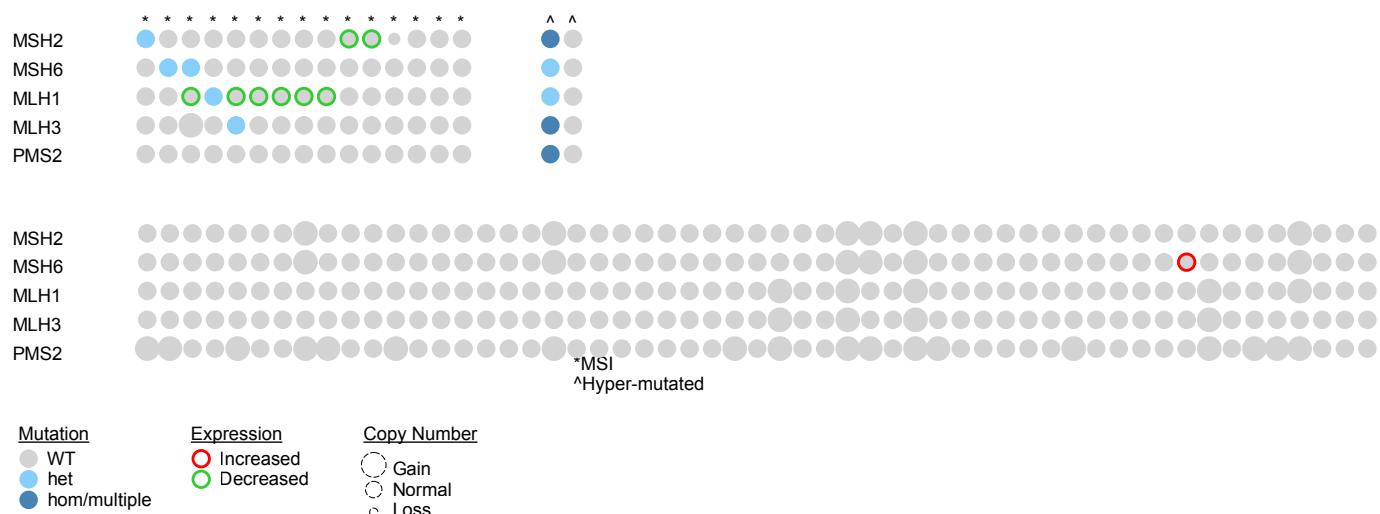
doi:10.1038/nature11282



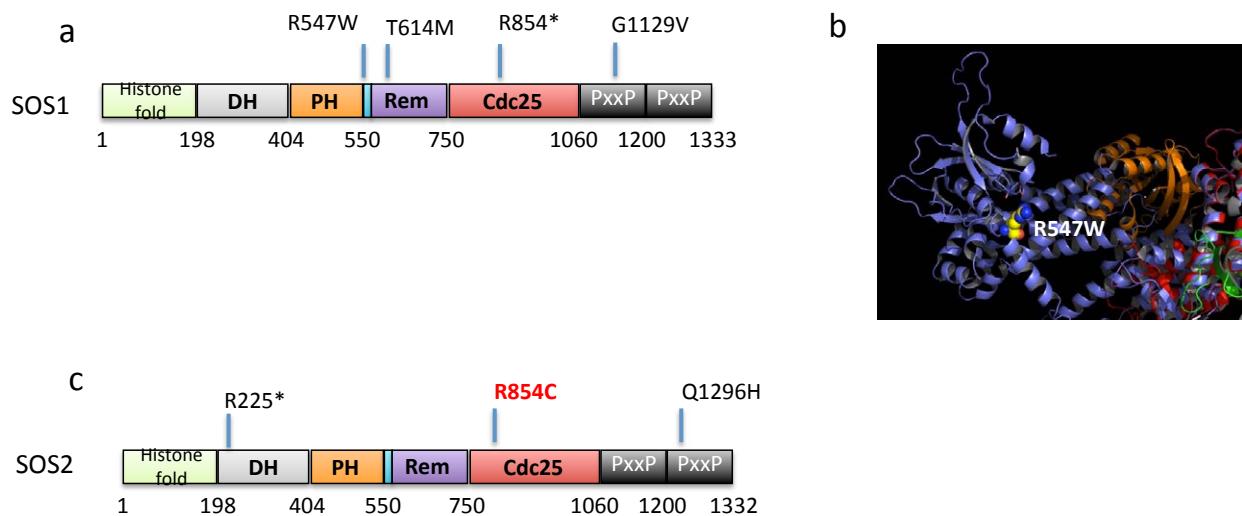
Supplementary Fig 1. Heatmap depicting exome coverage across samples and targeted CCDS regions. The red line below area of low coverage indicated off-target regions that were captured by the probe set



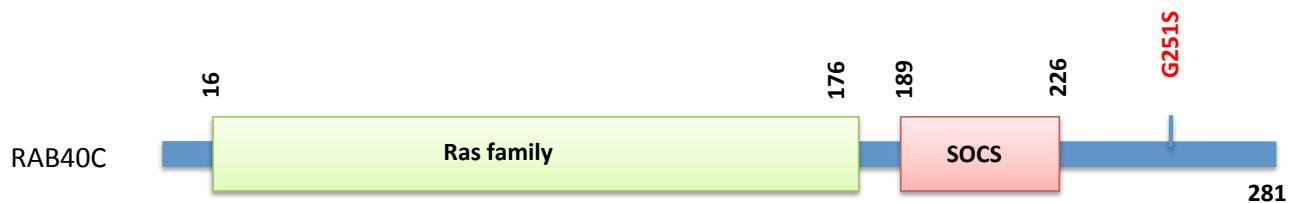
Supplementary Fig 2. Mutations that were predicted by exome sequencing were validated using the RNA-seq data. **(a)** Allelic expression counts for the mutant allele versus the reference allele in the tumor samples. A large fraction of the mutations could be validated by RNA-seq as the mutant allele was well expressed. **(b)** Normal samples do not show the expression of the mutant alleles



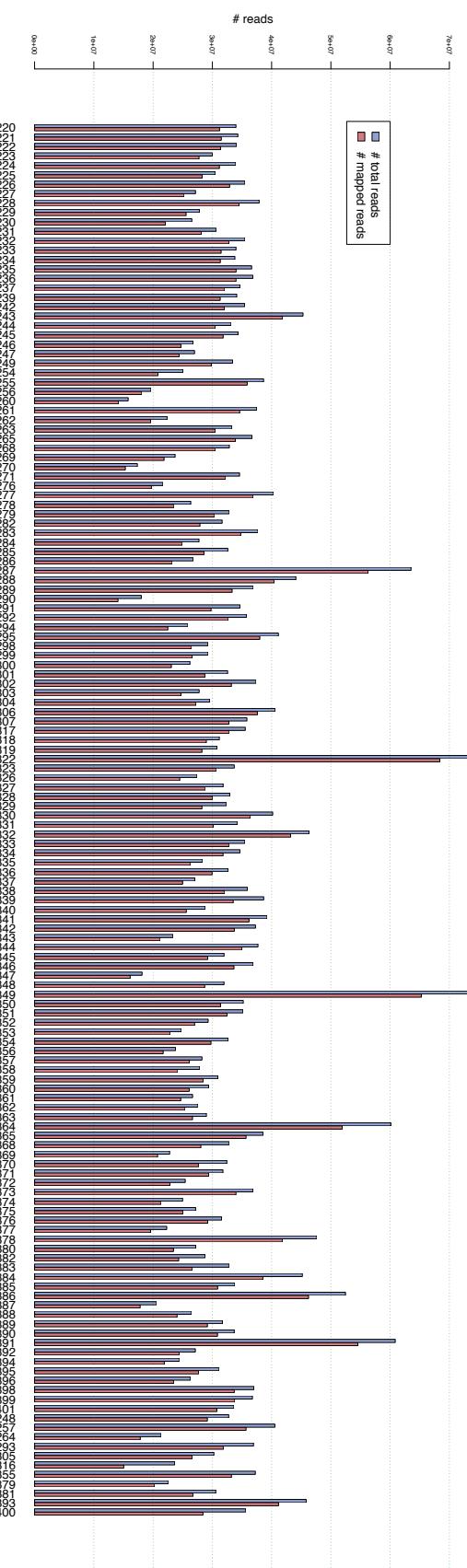
Supplementary Fig 3. Quilt plot depicting the alterations (mutation, copy number and expression) in mismatch repair genes in colon cancer samples. Each column of circles represents data for an individual tumor sample.



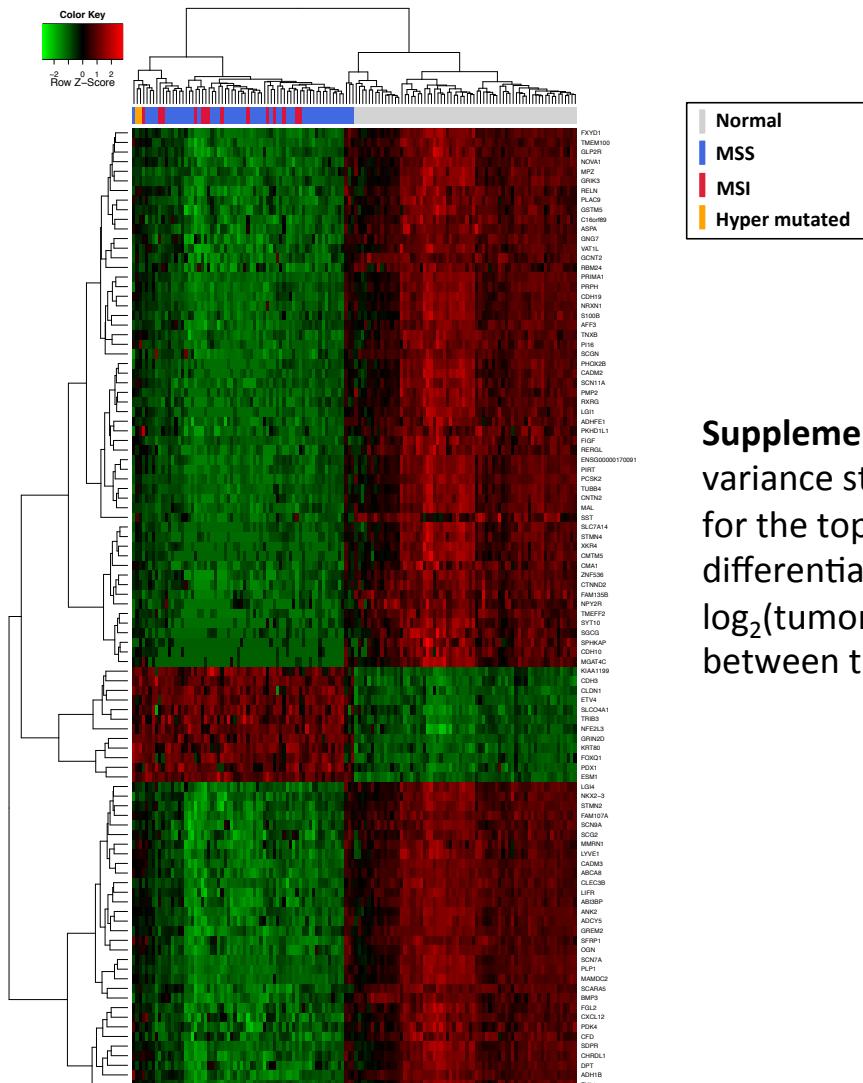
Supplementary Fig 4. Mutations identified in *SOS1* and *SOS2*. **(a)** *SOS1* mutations identified in this study shown on the domains of *SOS1*. **(b)** The R547W mutation mapped on to *SOS1* structure (PDB 1NVX) shows that it occurs in the helix linker that holds PH and DH domain and therefore this mutation may disrupt the interaction leading to *SOS1* activation. **(c)** *SOS2* mutations shown on *SOS2* domains. The hotspot mutations are shown in red and stars represent stop codons. DH- Dbl homology domain; PH - plekstrin homology domain; Rem - RAS-exchange motif; cdc25 – cdc25 domain; PxxP –polyproline domain.



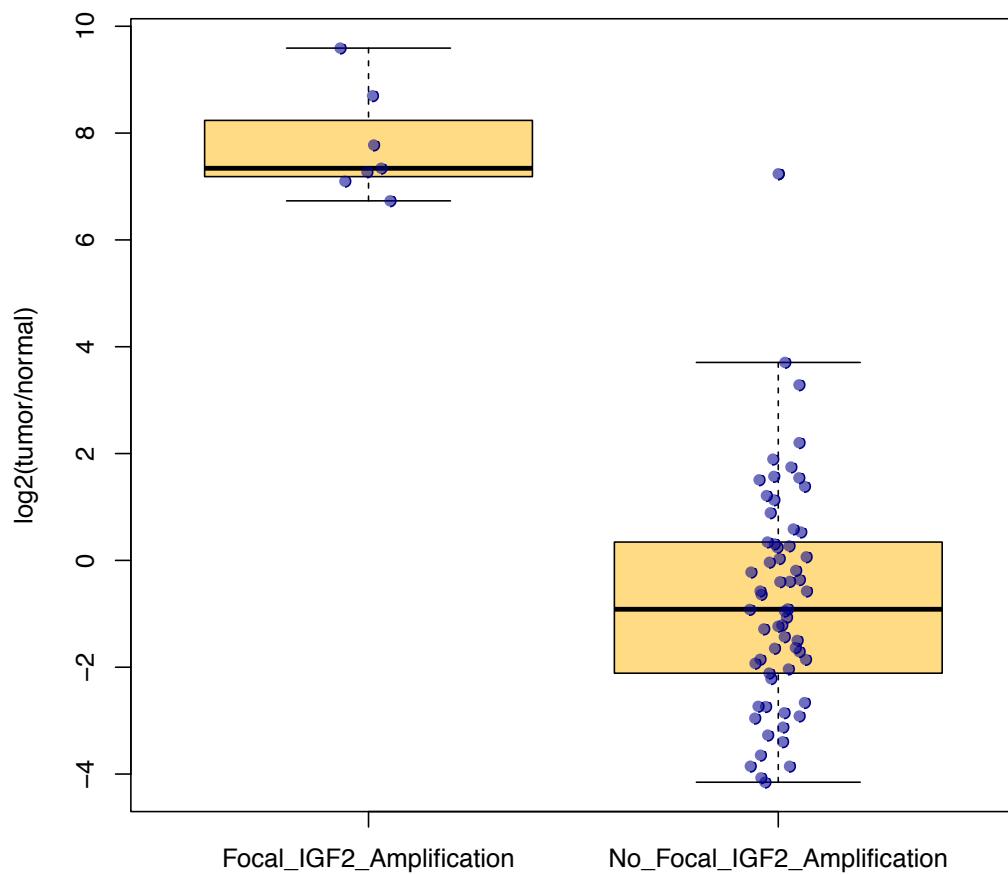
Supplementary Fig 5. Schematic diagram of *RAB40C* with recurrent mutations shown in red. SOCS - suppressors of cytokine signaling



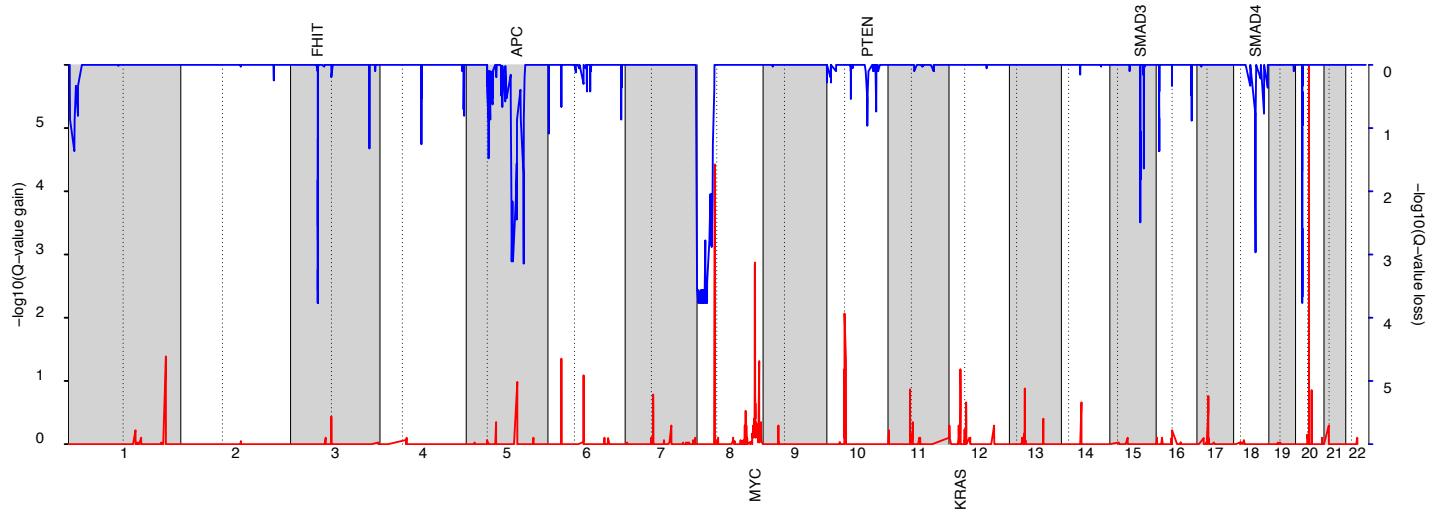
Supplementary Fig 6. Number of RNA-seq reads obtained and mapped by sample.



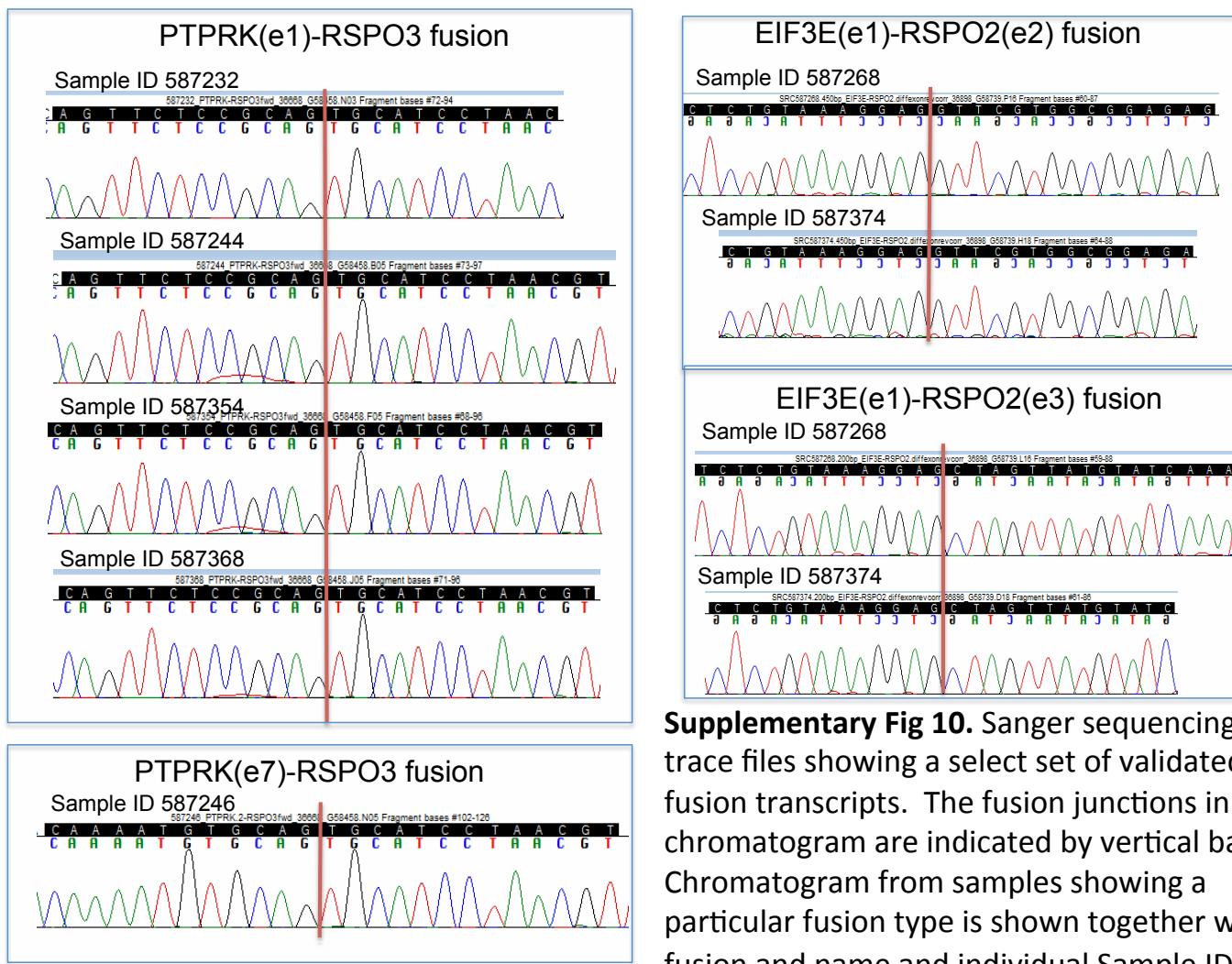
Supplementary Fig 7. Heatmap of the variance stabilized expression values for the top one hundred most differentially expressed genes with a $\log_2(\text{tumor/normal})$ ratio ≥ 3 or ≤ -3 between tumor and normal samples.



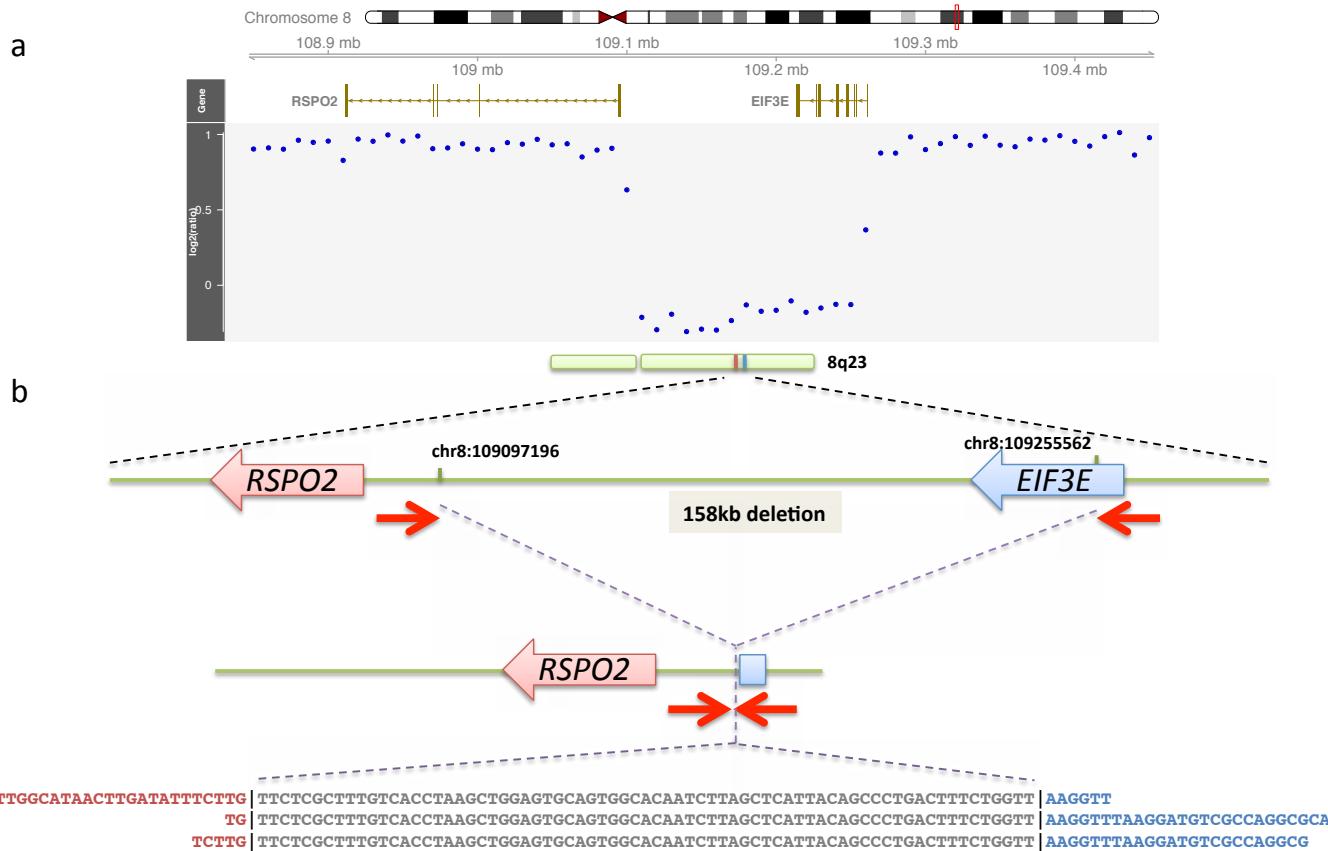
Supplementary Fig 8. Differential expression of *IGF2* plotted as $\log_2(\text{tumor/normal})$ values, in samples with amplification of the *IGF2* locus compared to samples without this amplification. *IGF2* is strongly upregulated in the samples with *IGF2* amplification.



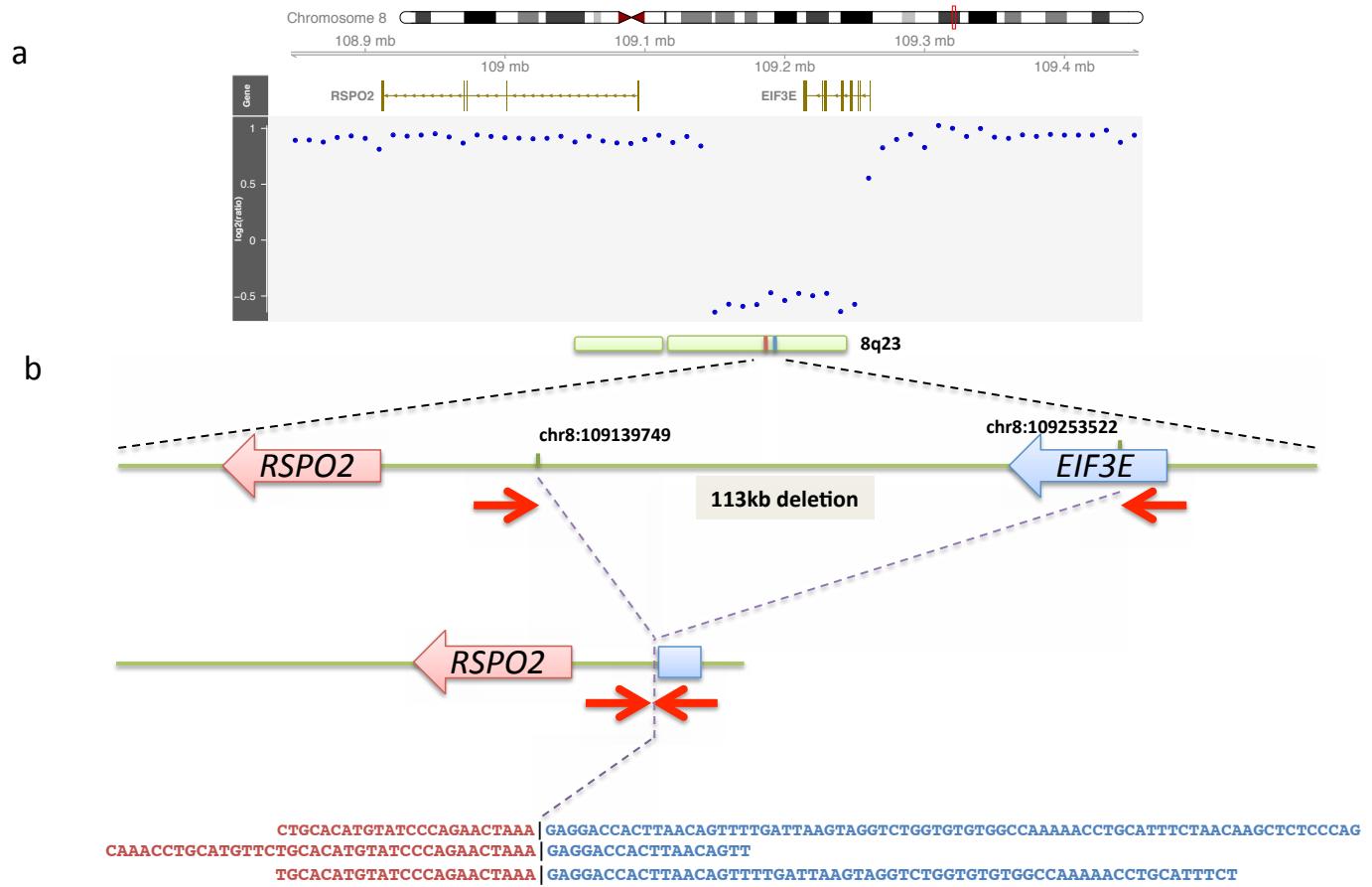
Supplementary Fig 9. GISTIC plot showing regions of recurrent loss (blue) and regions of recurrent gain (red). 18q is frequently lost, 8q and 20q are frequently gained. Genes of importance are annotated in the bottom (gains) and top (losses) of the plot.



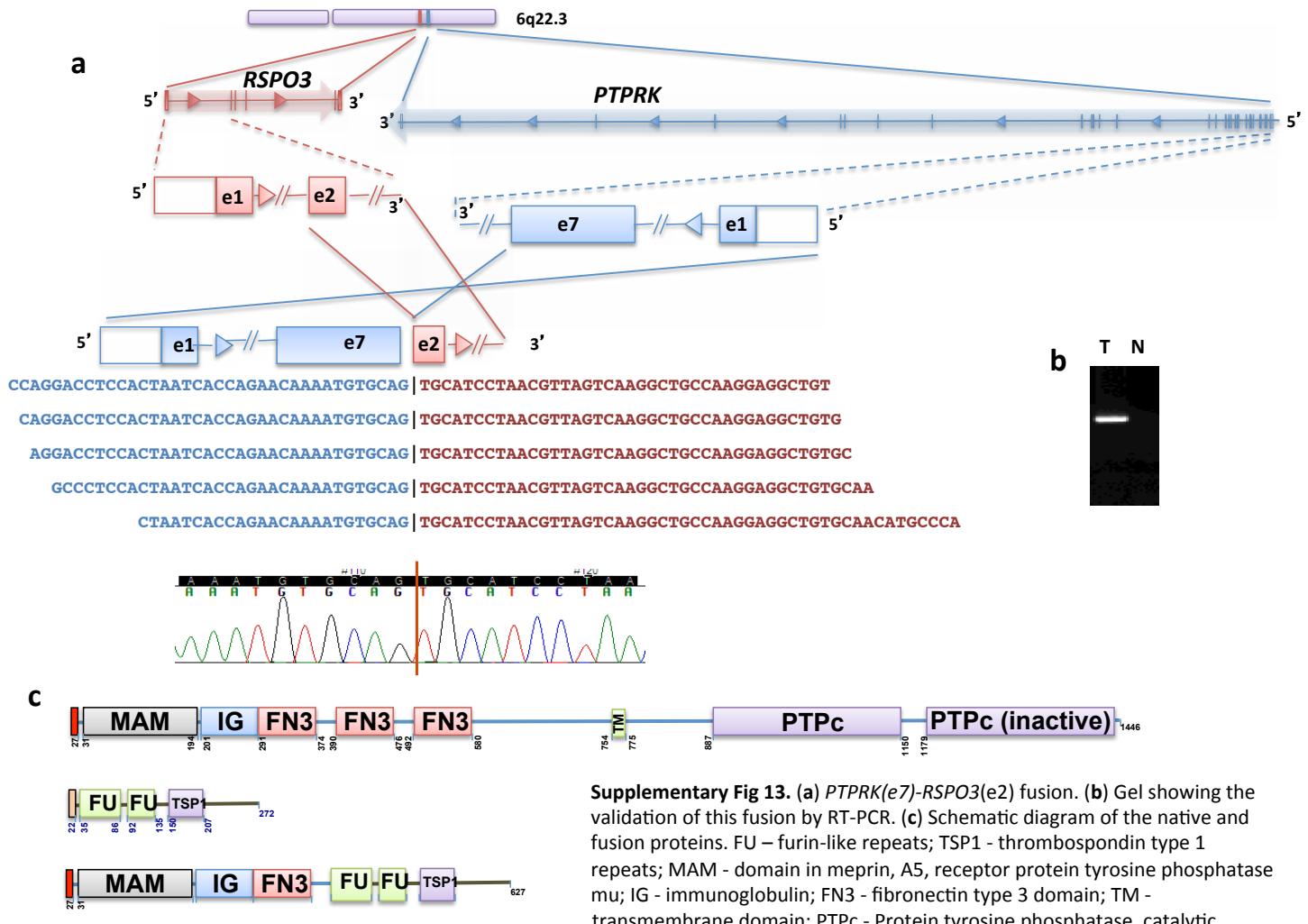
Supplementary Fig 10. Sanger sequencing trace files showing a select set of validated fusion transcripts. The fusion junctions in chromatogram are indicated by vertical bars. Chromatogram from samples showing a particular fusion type is shown together with fusion and name and individual Sample ID.



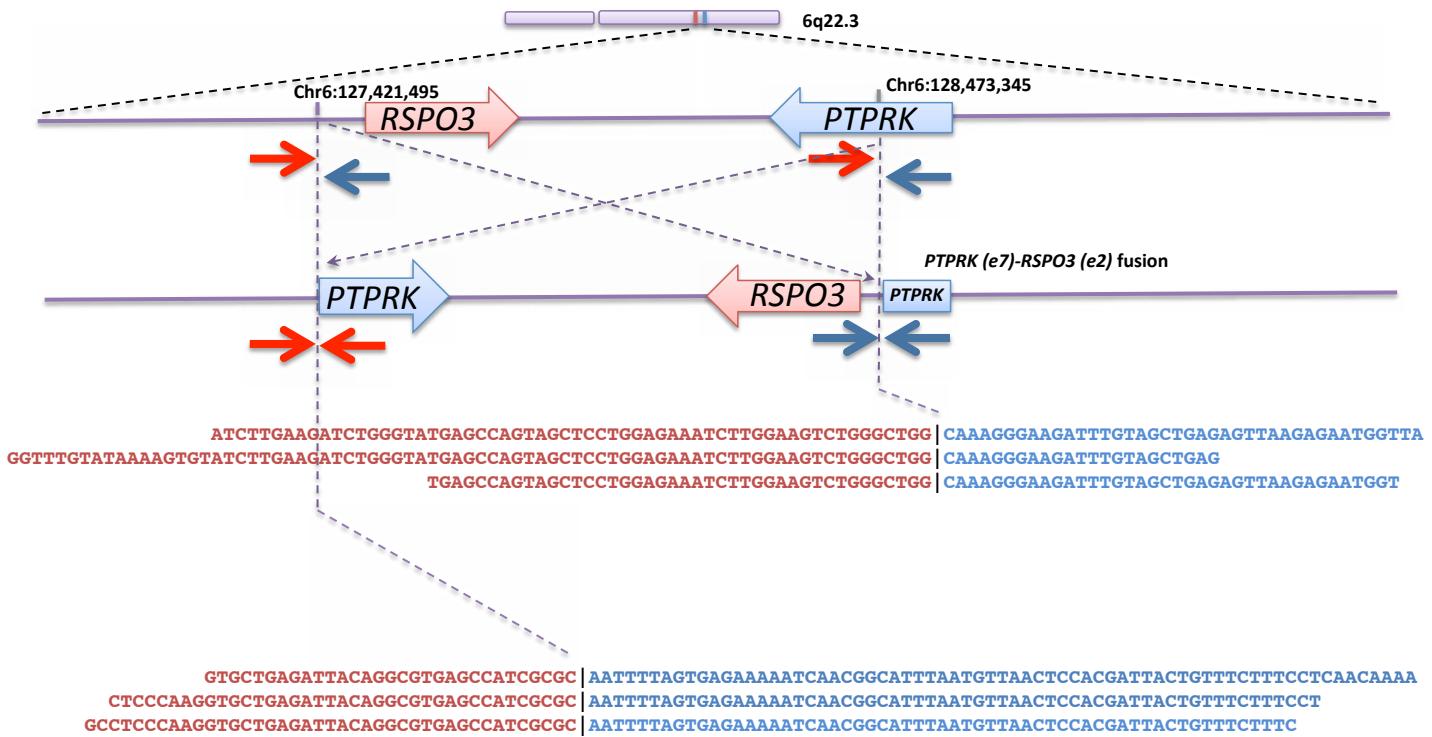
Supplementary Fig 11. (a) Copy number data computed from whole genome read counts binned over 10kb non-overlapping bins revealed that the genomic region containing EIF3E and RSPO2 is amplified in the tumors. However, within this amplicon the sequence between exon 1 of EIF3E and the proximal part of RSPO2 is deleted. (b) Schematic representation of the deletion in the genomic region containing EIF3E and 5' end of RSPO2 deduced based on the junction spanning (25) and mate-pair (10) reads. A representative mate-pair, shown as red arrows, when mapped to the tumor genome align adjacent to each other, though the sequences represented by them when mapped to the reference genome is separated by 158kb indicating that this region has been deleted in the tumor. A small insertion (sequence shown in grey) is observed in place of the deleted sequence. This represents data derived from whole genome sequencing of tumor sample (ID 587374) where EIF3E(e1)-RSPO2(e2) fusion was identified.



Supplementary Fig 12. (a) Copy number data computed from whole genome read counts binned over 10kb non-overlapping bins revealed that the genomic region containing EIF3E and RSPO2 is amplified in the tumors. However, within this amplicon the sequence between exon 1 of EIF3E and the proximal part of RSPO2 is deleted. (b) Schematic representation of the deletion in the genomic region containing EIF3E and 5' end of RSPO2 deduced based on the junction spanning (39) and mate-pair (54) reads. A representative mate-pair, shown as red arrows, when mapped to the tumor genome align adjacent to each other, though the sequences represented by them when mapped to the reference genome is separated by 113kb indicating that this region has been deleted in the tumor. This represents data derived from whole genome sequencing of tumor sample (ID 587268) where EIF3E(e1)-RSPO2(e2) fusion was identified.

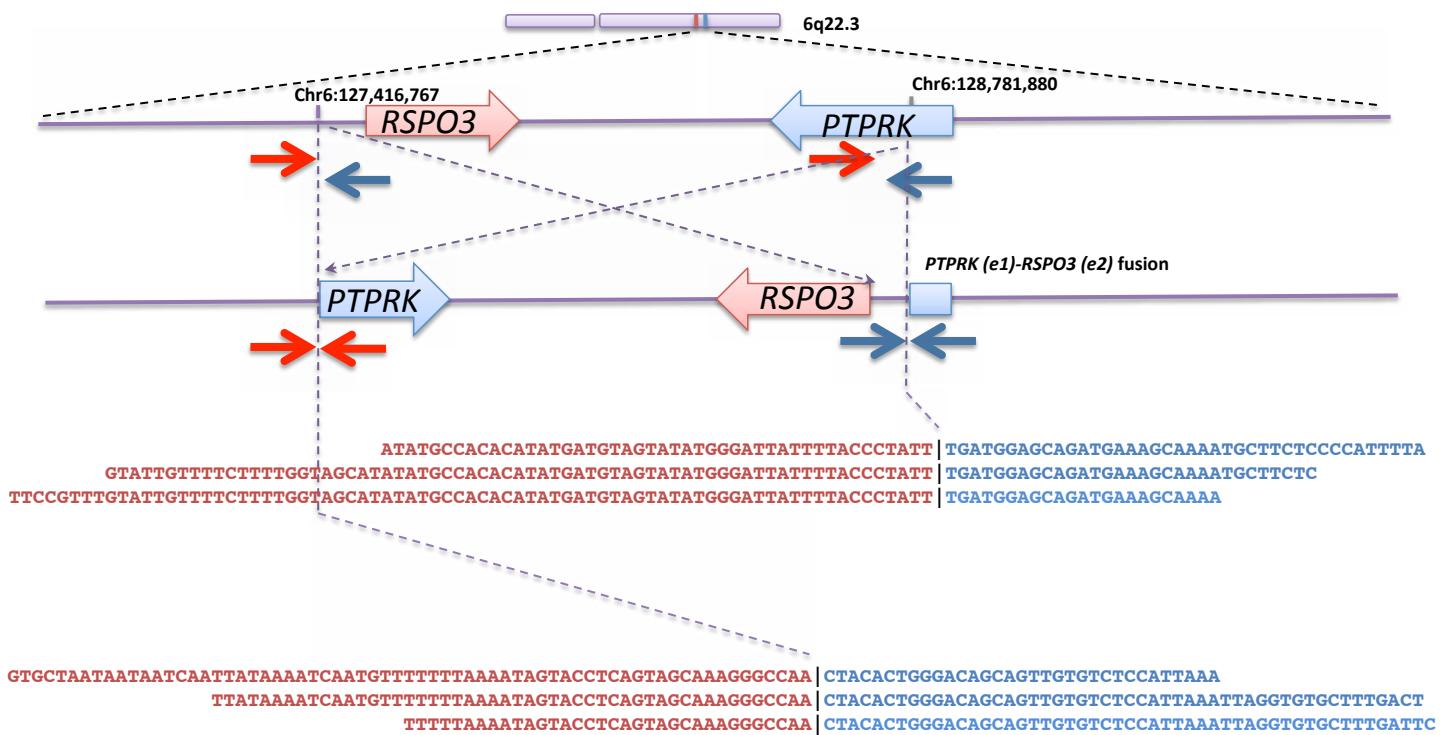


Supplementary Fig 13. (a) *PTPRK(e7)-RSPO3(e2)* fusion. (b) Gel showing the validation of this fusion by RT-PCR. (c) Schematic diagram of the native and fusion proteins. FU – furin-like repeats; TSP1 - thrombospondin type 1 repeats; MAM - domain in meprin, A5, receptor protein tyrosine phosphatase mu; IG - immunoglobulin; FN3 - fibronectin type 3 domain; TM - transmembrane domain; PTPc - Protein tyrosine phosphatase, catalytic domain.

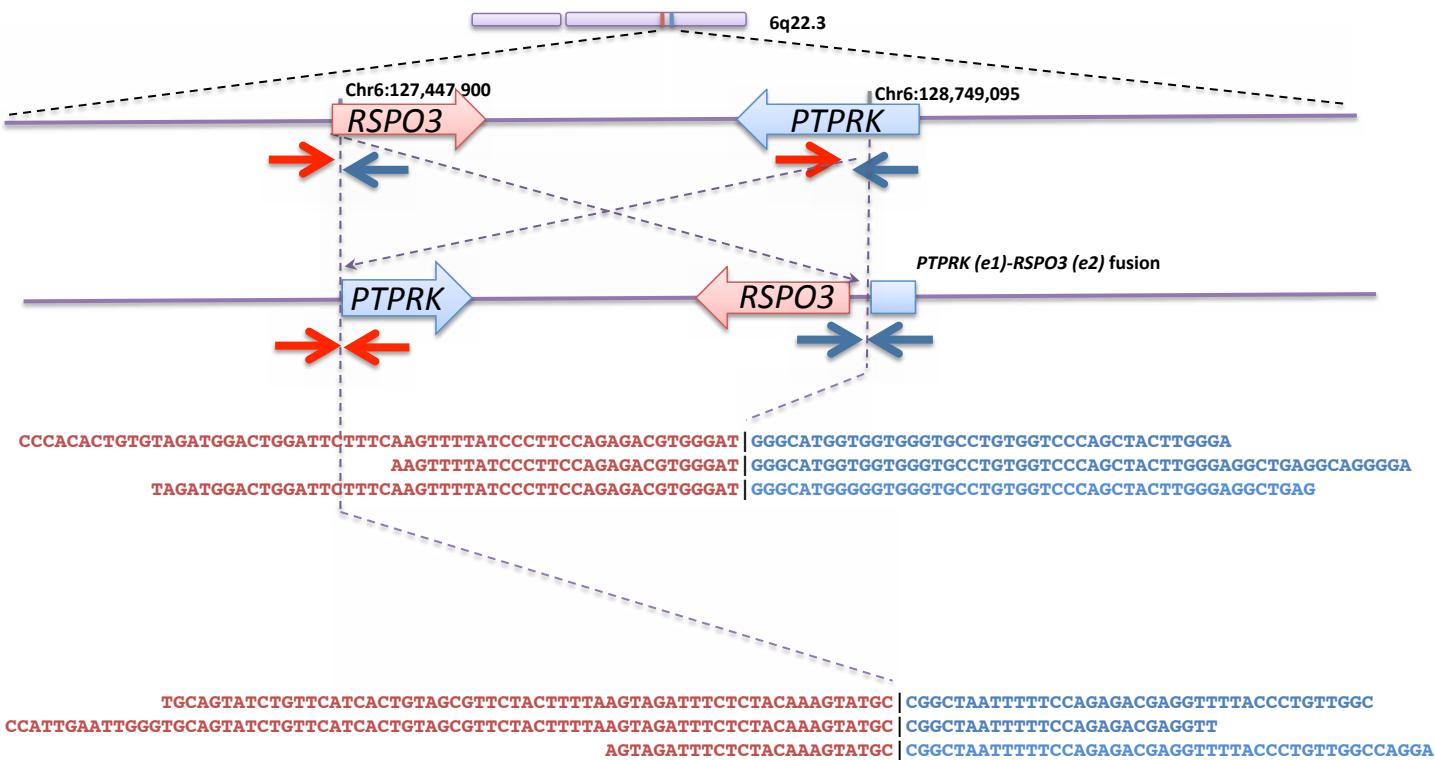


Supplementary Fig 14. Schematic representation of the inversion event in the genomic region containing *PTPRK* and *RSPO3* supported by junction spanning (15; 11) and mate-pair reads (2; 5) whole genome sequencing reads.

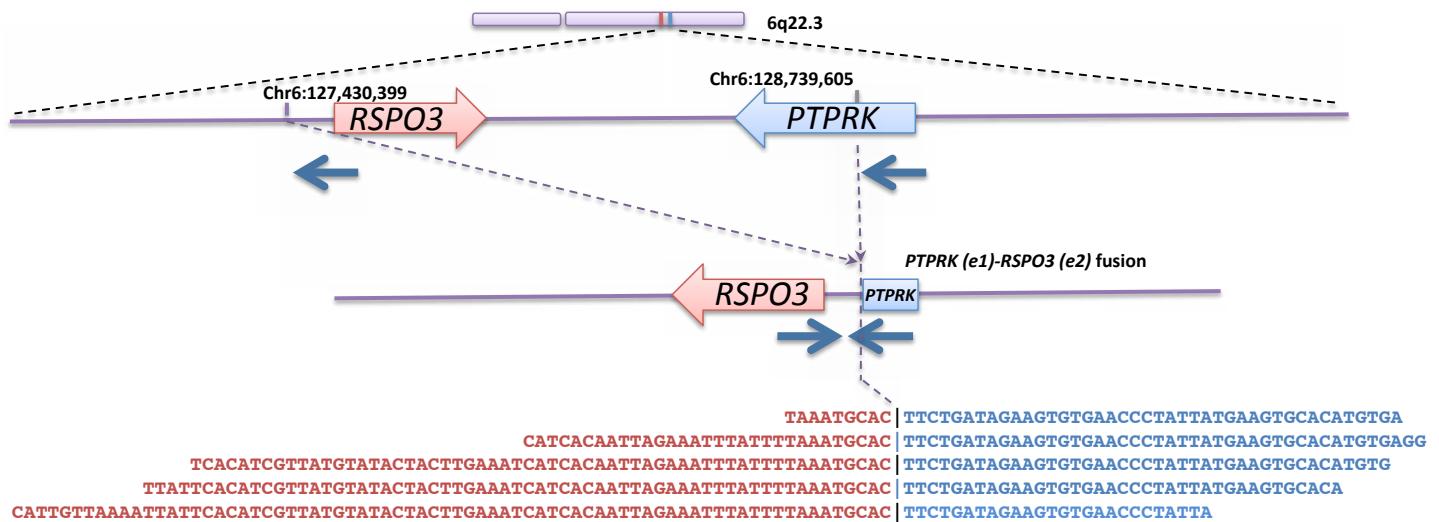
Representative mate-pairs shown as red or blue arrows when mapped to the tumor genome are adjacent to each other in opposition (point to each other). However, in the normal genome they map over distances longer than the average insert size (250bp) of the library constructed for whole genome sequencing in the same direction (do not point to each other). This represents data derived from whole genome sequencing of tumor sample (ID 587246) where *PTPRK*(e7)-*RSPO2*(e2) was identified.



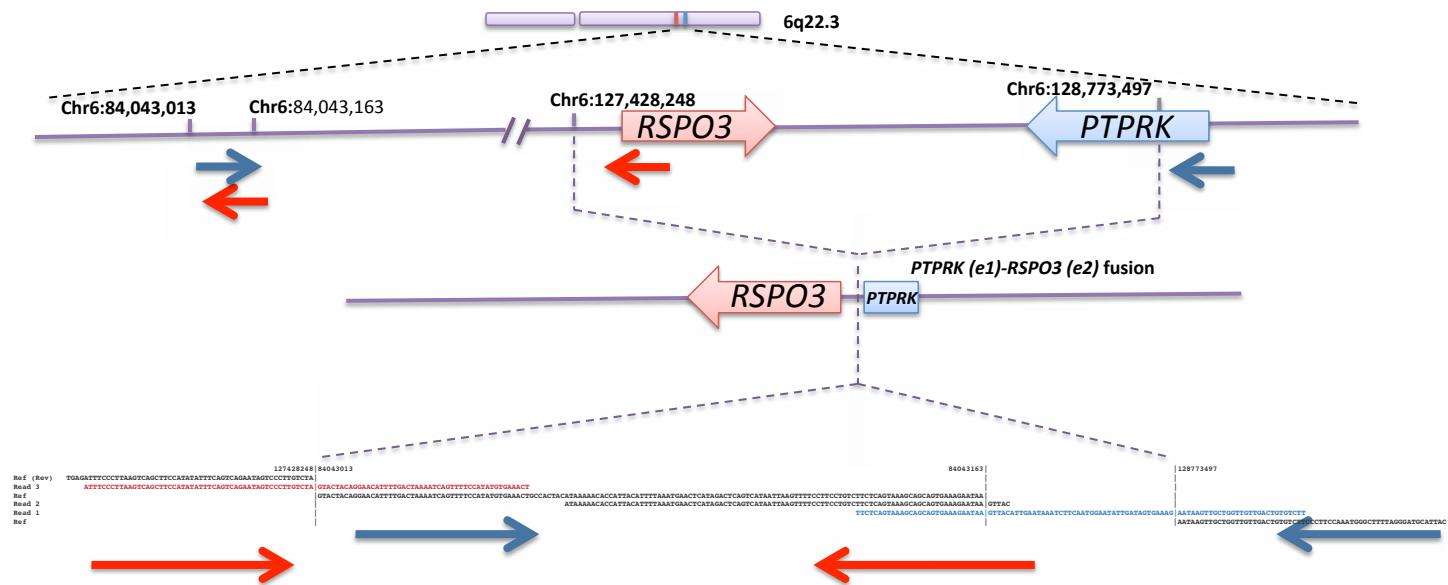
Supplementary Fig 15. Schematic representation of the inversion event in the genomic containing *PTPRK* and *RSPO3* supported by the junction spanning (18; 17) and paired-end (12; 13) whole genome sequencing reads. Representative mate-pairs shown as red or blue arrows when mapped to the tumor genome are adjacent to each other in opposition (point to each other). However, in the normal genome they map in the same direction (do not point to each other) over distances longer than the average inserts size (250bp) of the library constructed for whole genome sequencing. This represents data derived from whole genome sequencing of tumor sample (ID 587354) where *PTPRK*(e1)-*RSPO3*(e2) was identified.



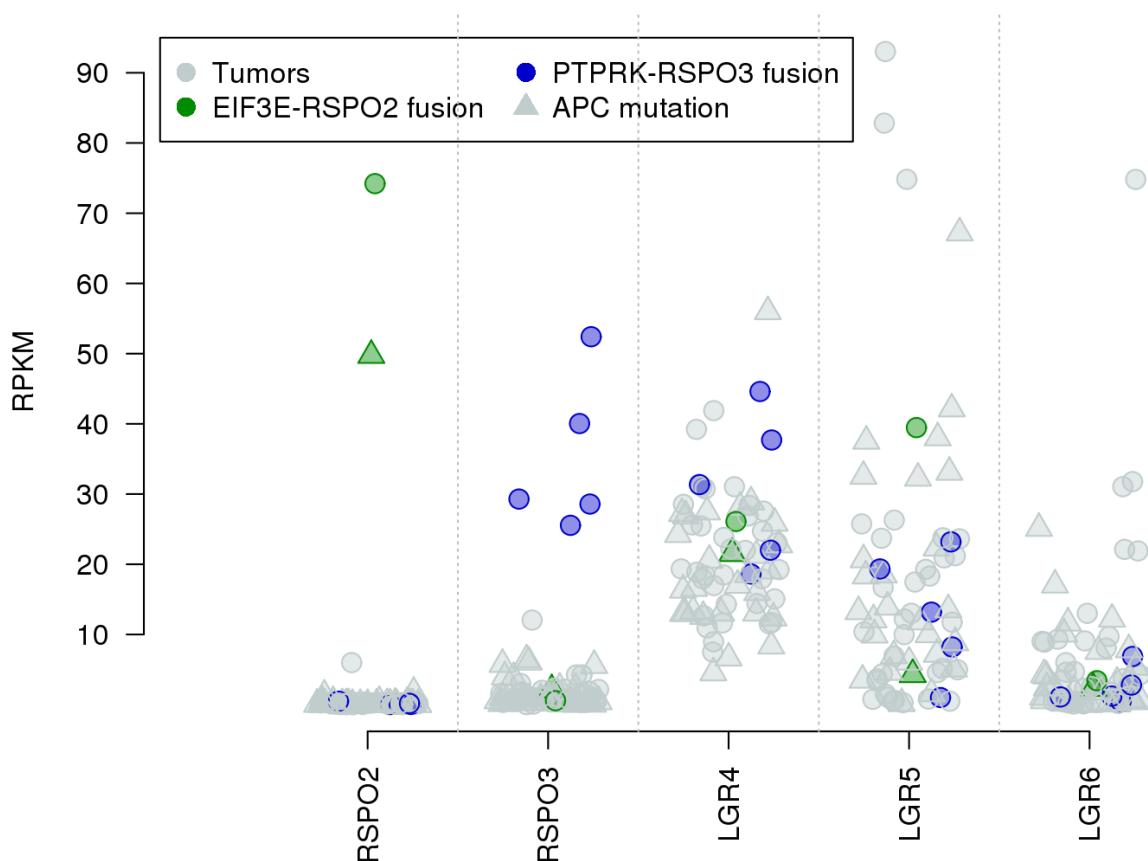
Supplementary Fig 16. Schematic representation of the inversion event in the genomic region containing *PTPRK* and *RSPO3* supported by the junction spanning (7; 6) and mate-pair (8; 7) whole genome sequencing reads. Representative mate-pairs shown as red or blue arrows when mapped to the tumor genome are adjacent to each other in opposition (point to each other). However, in the normal genome they map in the same direction (do not point to each other) over distances longer than the average inserts size (250bp) of the library constructed for whole genome sequencing. This represents data derived from whole genome sequencing of tumor sample (ID 587244) where *PTPRK*(e1)-*RSPO2*(e2) was identified.



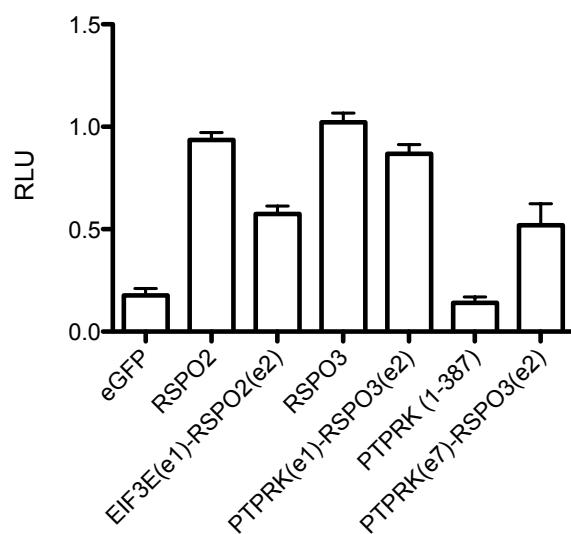
Supplementary Fig 17. Schematic representation of the inversion event in the genomic region containing *PTPRK* and *RSPO3* supported by the junction spanning (5) and mate-pair (4) whole genome sequencing reads. Representative mate-pairs shown as blue arrows when mapped to the tumor genome are adjacent to each other in opposition (point to each other). However, in the normal genome they map in the same direction (do not point to each other) over distances longer than the average inserts size (250bp) of the library constructed for whole genome sequencing. The inversion event based on sequence read evidence indicates that it is a complex inversion involving multiple break points resulting in an inversion. Only the configuration of the genomic region involving *PTPRK* and *RSPO3* is shown. This represents data derived from whole genome sequencing of tumor sample (ID 587368) where *PTPRK(e1)-RSPO3(e2)* was identified.



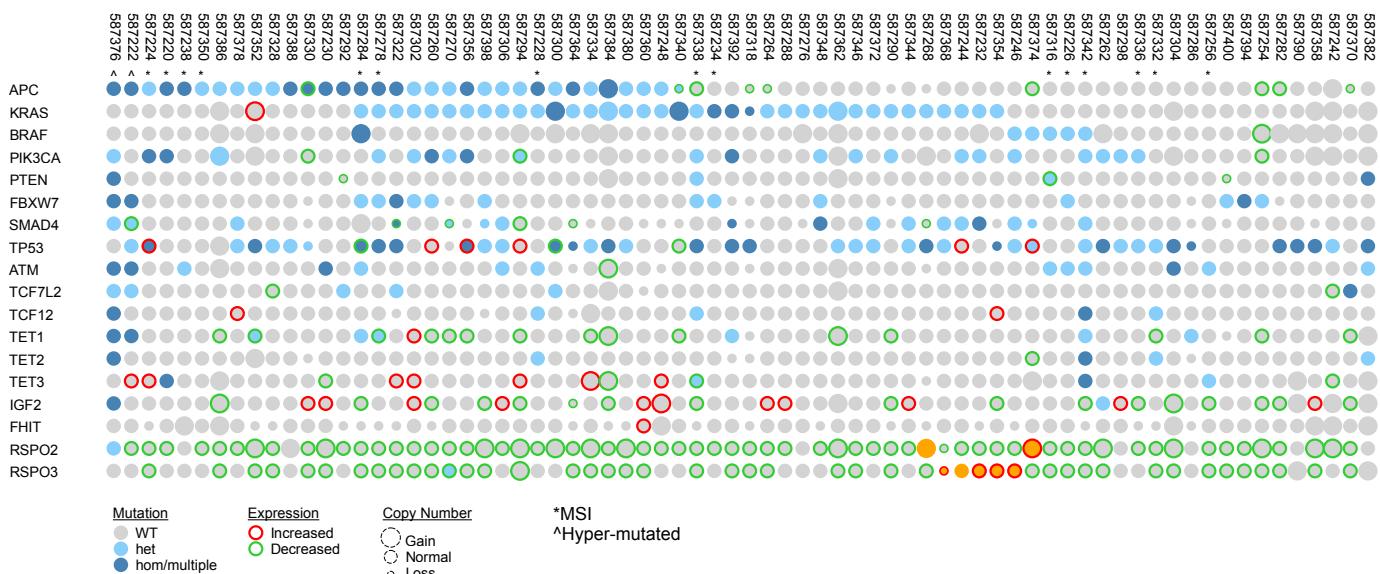
Supplementary Fig 18. Schematic representation of the inversion event in the genomic region containing *PTPRK* and *RSPO3* supported by the junction spanning and paired-end whole genome sequencing reads. Representative mate-pairs shown as red or blue arrows when mapped to the tumor genome are adjacent to each other in opposition (point to each other). However, in the normal genome the reads represented by the read arrow map in the same direction (do not point to each other) over distances longer than the average inserts size (250bp; reads represented by both blue and red mate-pair reads) of the library constructed for whole genome sequencing. The inversion event based on sequence read evidence indicates that it is a complex inversion involving multiple break points leading to the inversion. Only the configuration of the genomic region involving *PTPRK* and *RSPO3* reconstructed from the read evidence in this region is shown. This represents data derived from whole genome sequencing of tumor sample (ID 587232) where *PTPRK*(e1)-*RSPO2*(e2) was identified.



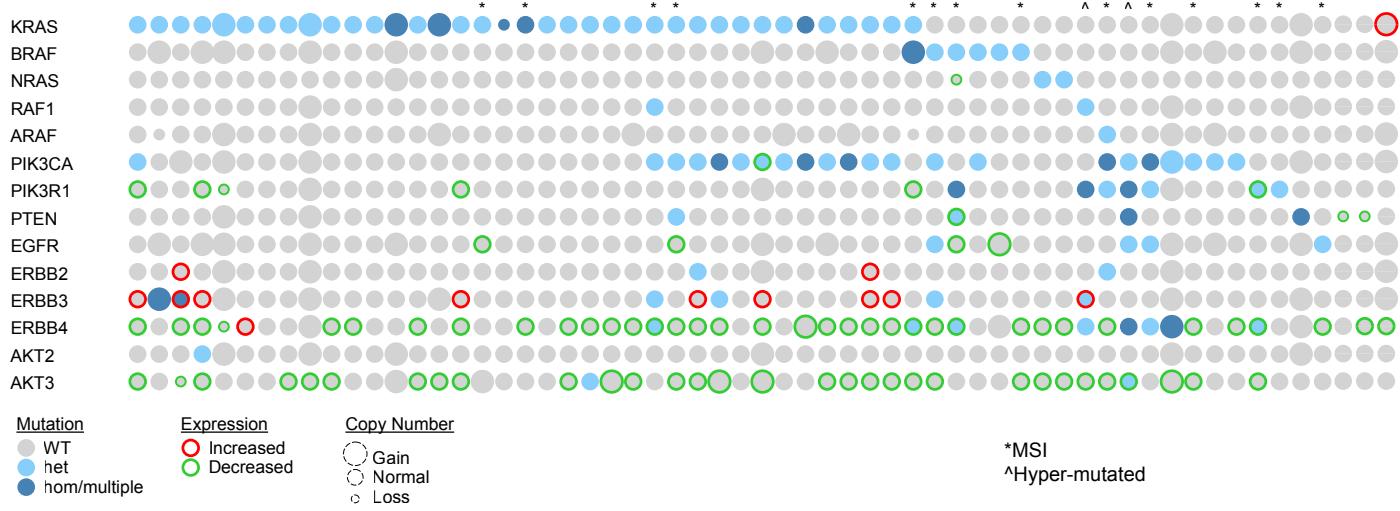
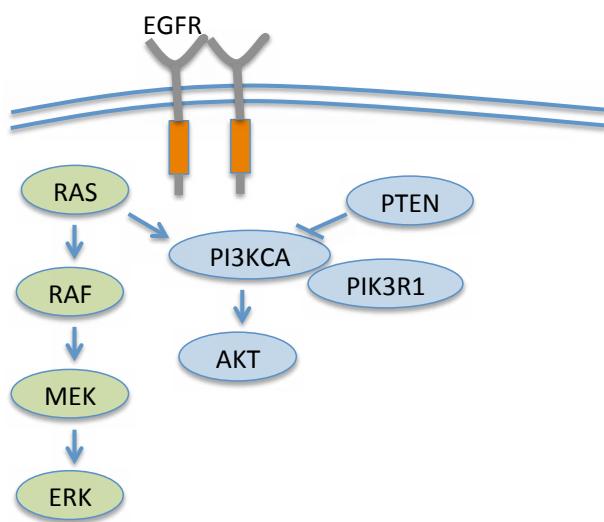
Supplementary Fig 19. Expression of RSPO2, RSPO3 and LGR family of genes in the tumor samples.



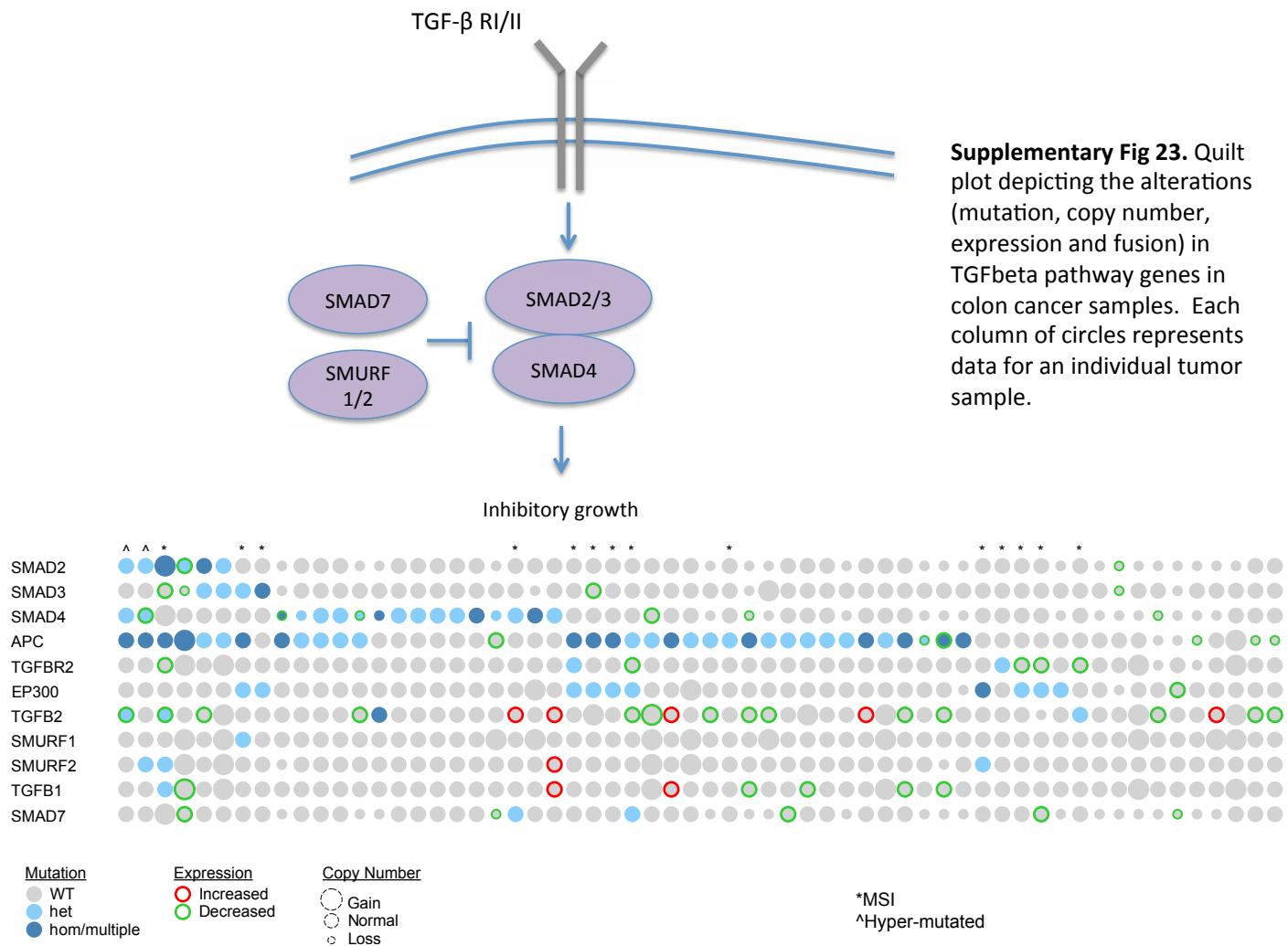
Supplementary Fig 20. Conditioned media derived from cells expressing the RSPO fusion constructs activate Wnt signaling in the HT-29 human colon cancer cell line. Though HT-29 carries mutations in each of its two APC alleles (E853* and E1554fs), the E1554fs mutant allele is functional (Oncogene 2001, 20:5025 and PNAS 2000, 97:3352) rendering HT-29 Wnt responsiveness. RSPO fusion proteins did not further activate the pathway in SW480 cells where both APC alleles (Oncogene 2001, 20:5025) are fully inactive (data not shown). Error bars represent mean \pm SD from three replicates.

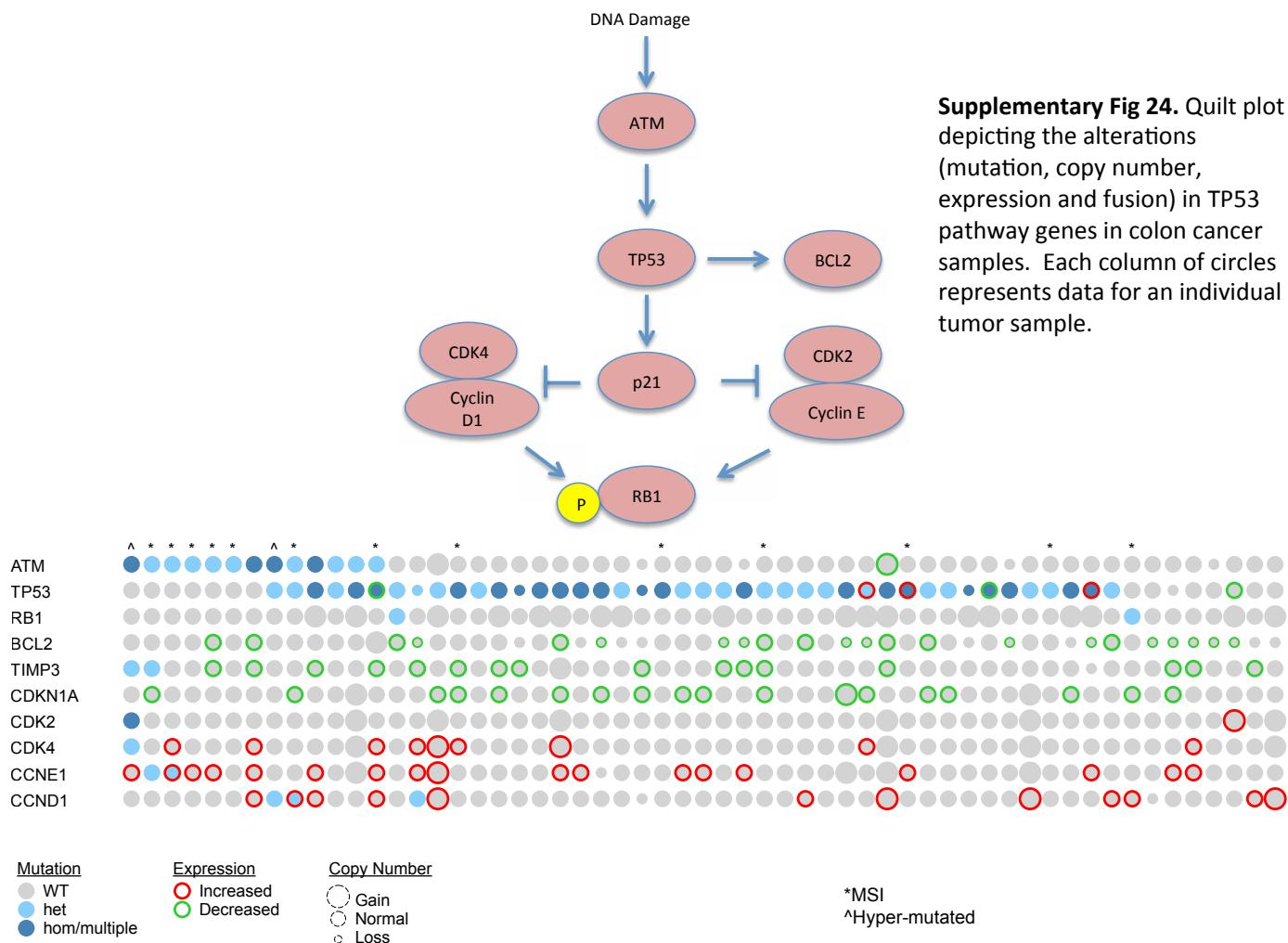


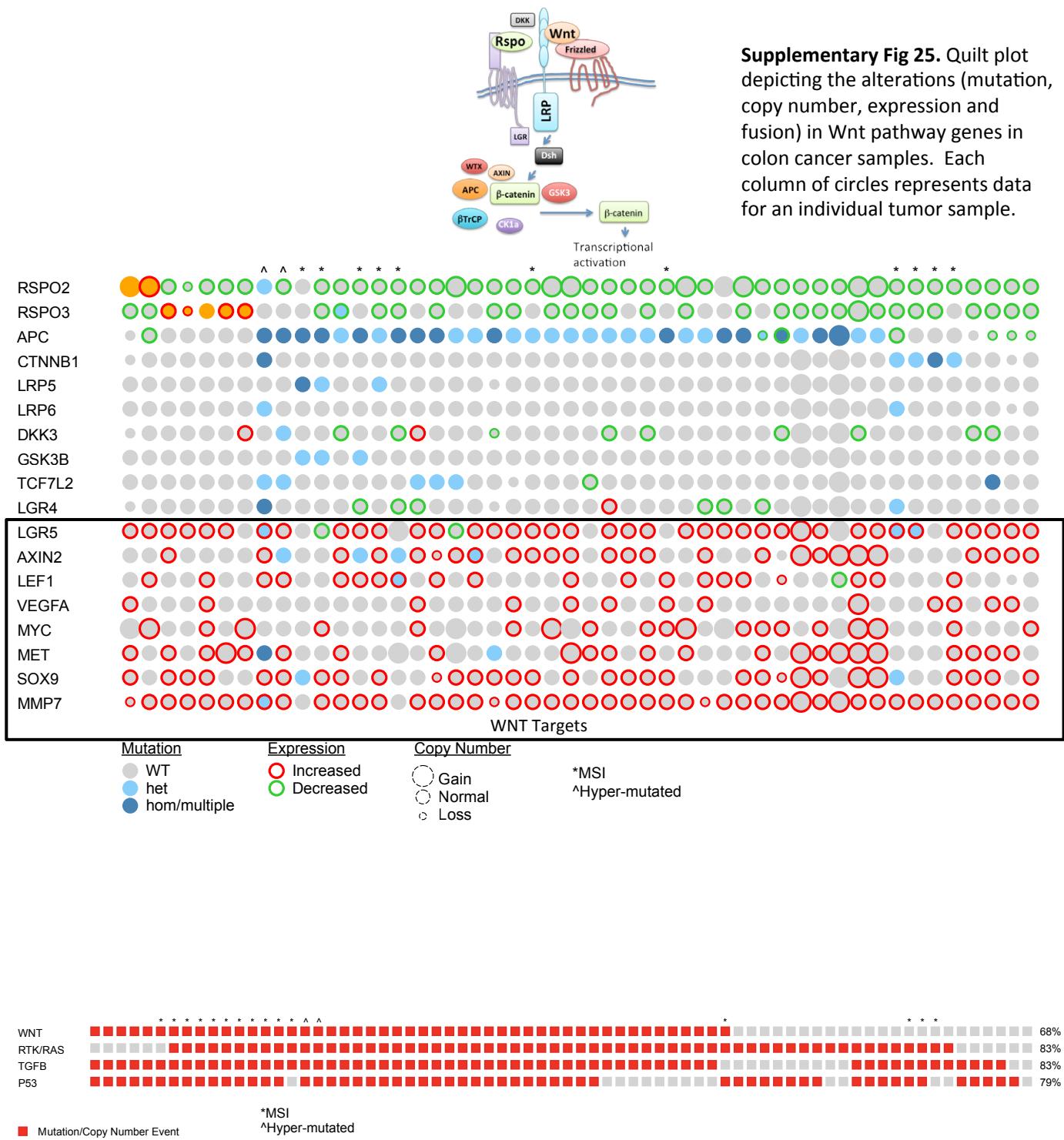
Supplementary Fig 21. Quilt plot depicting the alterations (mutation, copy number, expression and fusion) in select genes in colon cancer samples. Each column of circles represents data for an individual tumor sample.

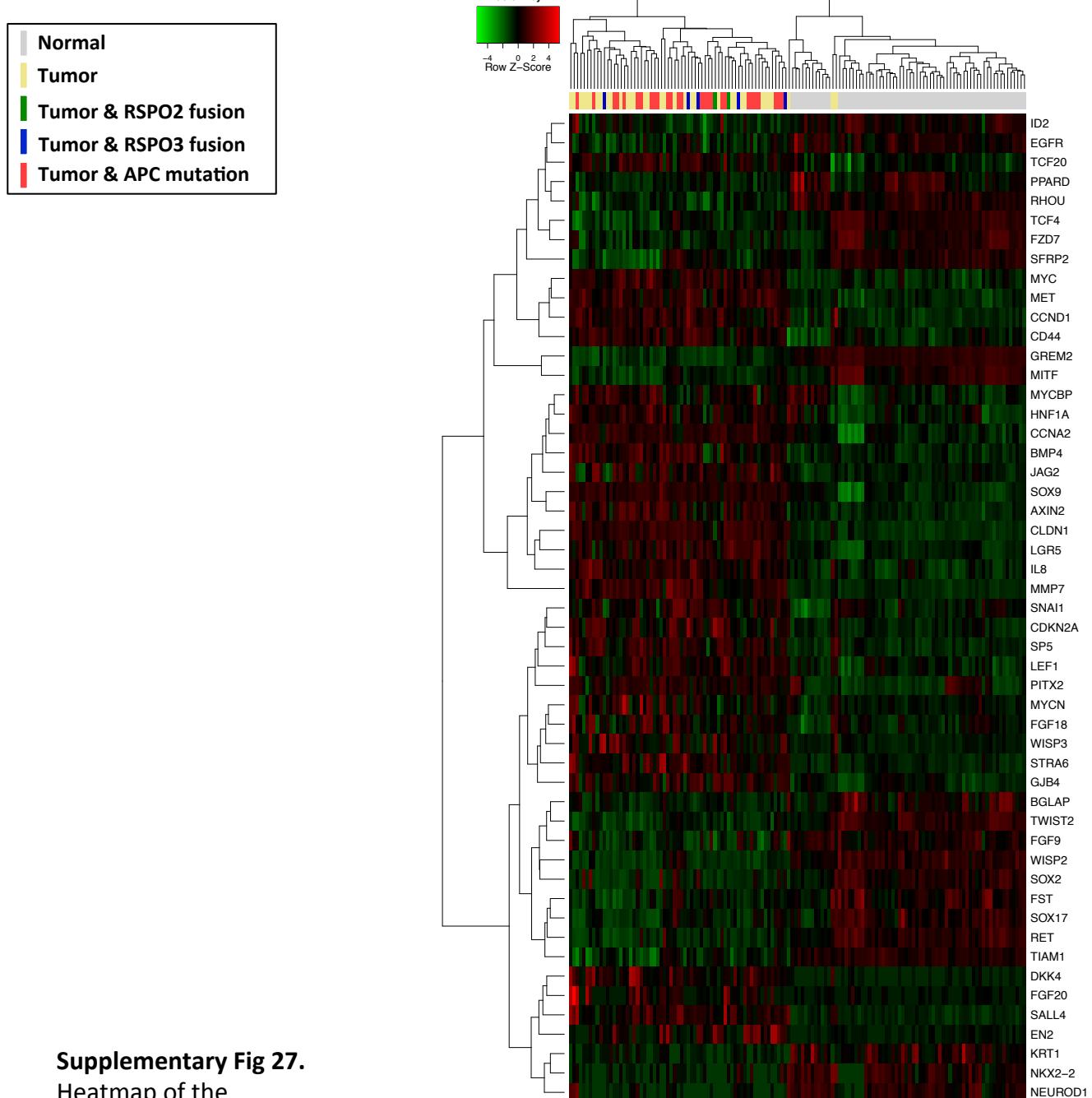


Supplementary Fig 22. Quilt plot depicting the alterations (mutation, copy number, expression and fusion) in RTK/RAS pathway genes in colon cancer samples. Each column of circles represents data for an individual tumor sample.



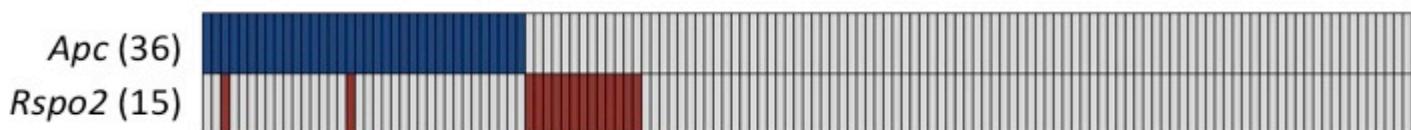




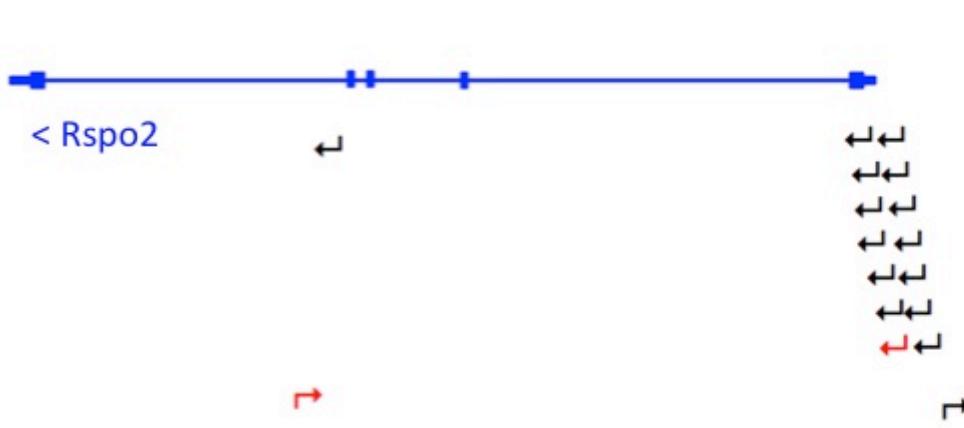


Supplementary Fig 27.
Heatmap of the variance stabilized expression values showing upregulation of multiple Wnt target genes both in *RSPO* fusion positive and *APC* mutant tumors compared to the adjacent normal samples.

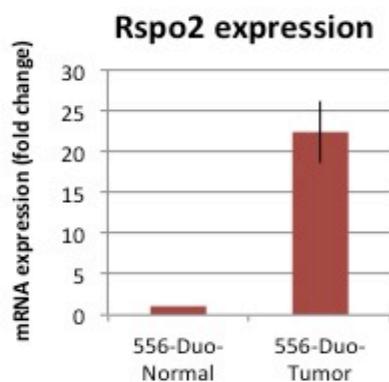
a



b



c



Supplementary Fig 28. *Rspo2* overexpression and *Apc* loss of function are mutually exclusive drivers of intestinal tumorigenesis in a murine model. A transposon-based genetic screen for intestinal tumors in mice was reported in Starr et al, 2009. Common insertion site (CIS) analysis identified *Apc* and *Rspo2* as candidate colorectal cancer genes. (a) Co-occurrence of transposon insertion sites in *Apc* and *Rspo2* show mutual exclusivity. Colored bars indicate tumors with transposon insertions in each gene. The total number (N) is indicated out of 135 tumors assessed. (b) Transposon insertion sites in the *Rspo2* gene. Arrows indicate the location of transposon integration and the orientation of the promoter-splice donor within the transposon. Red arrows indicate insertions in tumors with co-occurring insertions in *Apc*. The majority of insertions in this region are predicted to overexpress *Rspo2*. (c) Expression of *Rspo2* mRNA was evaluated by quantitative RT-PCR for a tumor with a transposon insertion near the *Rspo2* transcription start site, compared to adjacent normal tissue.